



戦略の自動進化および時系列乱数度による価格予測

田中 美栄子* · 楊 欣* · 糸井 良太*

Price Predictions Based on Evolvable Strategies, Principal Component Analysis, and Randomness of Data Strings

Mieko TANAKA-YAMAWAKI*, Xin YANG*, and Ryota ITOI*

Abstract– This paper summarizes the trials of predicting price movements in the financial market in various ways. The first example is to use the framework of the game theory between two players by regarding the next price movement as the next action of a player and the information of the environment surrounding the price movement as the action of the opponent player by using the evolutionary computation. The second example is to predict trendy sectors by using RMT-PCA. The third example is to use the randomness of the price fluctuations by means of the RMT-test. An empirical rule “High randomness predicts a good performance in the next period” is extracted by studying the randomness of stocks in the period of 2007-2009.

Keywords– evolvable strategies, game theory, RMT-PCA, RMT-test, randomness measure

1. はじめに

株価や為替などの金融商品の価格は大勢の投資家の意思と行動によって刻々と変化するため、これをモデル化したり、確率現象と考えて計算科学の対象にすることに意味があると考えられている。

そのような価格変動を決める要因は単一のものではなく、市場の様々な情報が複雑に交錯した結果であると思われる。しかも時々刻々とそれは変化するものである。仮に安定期に特定の要因が強く働くとしても状況に応じてそれは変化するに違いない。そう考えると何らかのドグマに基づいたモデルを先験的に仮定しても意味がなさそうに思われる。それよりも、客観的にどの期間にどの要因が働いたのかを過去データをもとに研究した後、モデルを構築し、その結果を現実と照らし合わせて検討するというアプローチを取るほうがより科学的だと考えられる。実は現象を観察してそれをモデル化し、コンピュータ・シミュレーションで再現する、というのは物理学における標準的な問題の解き方である。(人気俳優の福山雅治演じる湯川学博士が「解った!」と言って数式を書きなぐる時、彼は思いついたモデルを書き留めているのだと思う。)

対象が自然現象である場合は、短時間の間に相手の本質的な構造が大きく動くことがないため、それを観察する自分をモデルに組み込む必要はない。ところが社会現象となると、観察者自身の行動が全体を動かす要因ともなるため、単純な方程式で表すことは普通、不可能である。このように「相手もこちらと同様の知能を持ち、こちらが考えることは相手もやはり考える」という状況は19世紀から20世紀にかけて大成功した従来の自然科学の方法論だけでは立ち行かない[1]。

その考えに沿って、市場価格の短期予測をゲームシミュレーションの方法で扱うことを考える。単純な価格変動モデルを仮定して、進化的アルゴリズムによるエージェント・シミュレーションによって投資戦略を自動生成するプログラムを作り、実際の金融データに適用する。ここでは投資戦略を、金融市場という刻々と変化する環境に適応しながら進化するエージェントとみなし、他のエージェントの活動結果によって生じる市場価格の動きを相手として様々な戦略を実行し、結果の可否に応じてその時点での環境(価格変動)に対して最適な戦略に進化してゆくものとする。

最も簡単な環境設定として、1歩前の価格変化の符号の情報を採用し、この情報を元に1歩後の変動の符号を予測することから始め、合計7歩程度まで参照する履歴を増やして行くと、3~4歩位までは予測率が次第に向上して約70%の予測率に達するが、5歩位で飽和してしまいそれ以上は向上しないという結果になった[2]。ここ

*鳥取大学大学院工学研究科 鳥取県鳥取市湖山町南 4 丁目 101

*Department of Information and Electronics, Graduate School of Engineering, Tottori University, 4-101 Koyamacho-Minami, Tottori

Received: 22 July 2013, 8 September 2013

では短期データとして記録される最小単位であるティック (tick) を 1 歩と仮定するが、これは時間の単位ではなく、取引間の間隔であるため一定ではなく、人気株等では数秒と短い、取引の少ない株ならばもっと長い時間が対応する。次にもう少し環境設定を現実に近いものにするため、いくつかの「指標」を用いてこれらと現在価格との上下関係を利用する。ここでいう指標とはその時刻に投資家が考慮したいと考える数値のことで、例えば直近の 1 週間の株価の平均値、または何らかの数量から計算できる量のその時点での値などが使われることが多い。ここでは広く使用される 10 個の指標 [3] をベースに、これらの内からいくつかを進化計算により選んで使用することにより数分先位までの価格の上下を予測したところ、株価によっては 6 割を超える的中率が得られた [4, 5]。

しかしこれらは 1993 - 2002 年の NYSE (ニューヨーク証券取引所) で取引された株価のティックデータ [6] を用いて得られたものであり、2010 年以降に株式市場に取り入れられた arrowhead と呼ばれる極短期の取引間隔に対応したものではない。これについては 4 章でも触れるが、市場の根本原理が前の結果を元に次の手を考えるという思考回路が何らかの形で成立するシステムである限りは、高速計算機による計算で対応できるため、手法自体はそのままではないにしても、今後も応用価値があると考えられる。

2. 予測手法: 履歴 (H) と予測ゲノム (P)

ここでは、価格を予測しようとする主体が、自身の置かれた環境である市場の動きを相手にゲームを行う、という見方で問題設定を行うこととする。市場の価格変動をリアルタイムで扱うのは金融のプロ以外は難しくお金もかかるので、過去の価格データ時系列をゲームの相手とみなし、相手に勝てる戦略を進化的計算法により一定期間のデータで学習し、それを使用して学習後の一定期間に次々と予測を行った場合にどの程度の的中率を得られるか、という問題を解くのである。

最も簡単な環境設定として、1 歩前の価格変化の符号の情報を採用した場合、価格変動を価格の { 下降, 不変, 上昇 } の 3 通りとし、それぞれを {0, 1, 2} と符号化する。この符号を過去 N 歩にわたって記録したものを履歴 (H) とする。履歴 (H) の各要素 h_i には i 歩前の価格変動が記録される。つまり $h_i \in \{0, 1, 2\}$ の値が N 個並んだものとなる。

$$H = h_1 h_2 h_3 \dots h_{N-1} h_N \quad (1)$$

履歴 (H) の長さは N であるので、可能な履歴 (H) のパターンとしては 3^N 種類ある。このそれぞれに対応した予測を要素として持つ、長さ 3^N の予測ゲノム (P) を

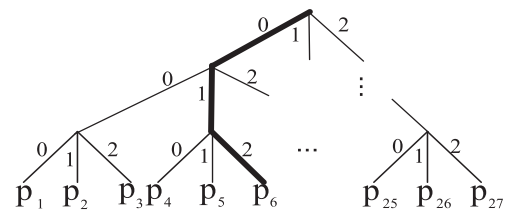


Fig. 1: 予測ゲノム (P) の木構造の例 ($N = 3$ の場合)

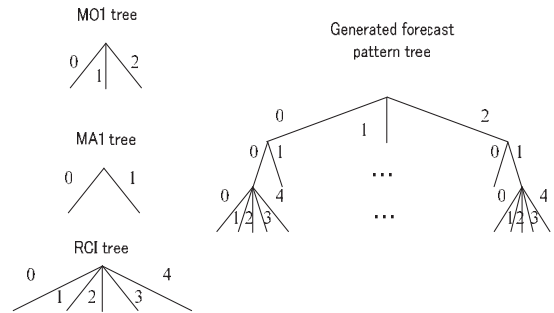


Fig. 2: 3 種の指標 (MO1, MA1, RCI) に対する木構造の例

進化計算で求める。予測ゲノム (P)

$$P = p_1 p_2 p_3 \dots p_{3^N-1} p_{3^N} \quad (2)$$

の要素 p_i には $\{0, 1, 2\}$ のいずれかの値が格納され、履歴 (H) のそれぞれのパターンに対応した次回の変動の予測を示す。 $N = 3$ の例を Fig. 1 に示す。長さが 3 の N 乗であるゲノムは高さが N の 3 分木の葉に対応させて表現できる。このとき仮に $H = 012$ であったとすれば、 $h_1 = 0$ であるので根では左の枝を選ぶ。同様に $h_2 = 1$ なので次は中央の枝を選ぶ。次は $h_3 = 2$ なので右の枝を選ぶ。よってこのときは p_6 の値が予測値となる [2]。初期条件では式 (2) の予測ゲノム (P) は $\{0, 1, 2\}$ を要素とするランダム列であり、100 個の異なるランダム列を使った予測に於いて、高い的中率を示した 10 個を優良なゲノムと見立てて次世代に残し、残りの 90 個は捨てる。その上で生き残った優秀な 10 個から其々 1 文字をランダムに変化させて亜種を 90 個作り、これを次世代の 100 個のゲノムとする。この一連の操作を一代として、世代交代を行い、優秀な遺伝子を自動生成する。

以上は履歴として 1 歩前の価格のみを仮定した場合であるが、履歴に使う環境要因を増やすと木構造は複雑になる。例えば MO1 (モメンタム = 現在価格と 1 歩前との比較を示す)、MA1 (単純移動平均)、RCI の 3 種類の指標を使った履歴を考える場合は Fig. 2 のような木構造を考えねばならない [4, 5, 7]。

この場合問題となるのは、どの指標をいくつ選択すべきかということも環境次第なので、環境の変化に合わせ

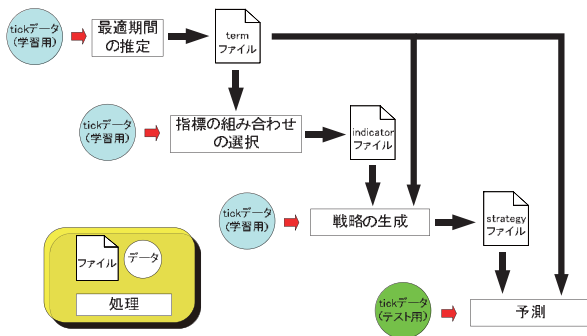


Fig. 3: 提案する予測プログラムの流れ

Table 1: 株価 8 銘柄に対する 10 歩後の予測的中率 (1993) [4, 5, 7]

| 銘柄(業種) | 予測的中率(%) | | データ数 | Tick 間隔 (s/tick) |
|----------|----------|-------|--------|------------------|
| | 最良 | 平均 | | |
| SMRT(小売) | 67.15 | 64.02 | 12525 | 473 |
| APC(石油) | 57.35 | 55.23 | 23685 | 253 |
| BBY(小売) | 58.79 | 57.34 | 54821 | 109 |
| CA(計算機) | 63.03 | 61.77 | 65051 | 92 |
| BP(石油) | 65.36 | 63.44 | 73562 | 83 |
| F(車) | 68.70 | 67.75 | 194561 | 32 |
| GM(車) | 78.54 | 77.37 | 277241 | 23 |
| IBM(計算機) | 82.25 | 81.38 | 455233 | 14 |

で最適な選択をしなければならないということである。そこで、進化計算で自動選択できるよう、ゲノム長を可変にしておく。また MA1 のように移動平均を使用する場合は、平均を取る区間 n の最適値も自動選択できるように設定しておく必要がある。このシステムの仕様を Fig. 3 に示す [5]。

3. 複数の指数を組み合わせた予測結果

NYSE の株価 [6] を 4 業種から 2 銘柄ずつ、合計 8 銘柄 (APC, BBY, BP, CA, F, GM, IBM, SMRT) を選び、これらの株価について、1993 年の前半を学習用データとし、後半をテストデータとして 10 歩後の価格の上下を予測した。結果として得られた予測的中率を Table 1 に示す。

IBM, GM 株は的中率が 80% 程度とかなり高いが、APC, BBY 株は 57% 程度とかなり差が見られる。これは、IBM や GM といった人気株はデータ数が多く一歩の間隔が短いことが原因であると考えられる。つまり、データ数の多い銘柄ほど近い未来を予測して的中率が高くなっていると考えられる。

次にどの指標が有効かを評価するため、異なる指標

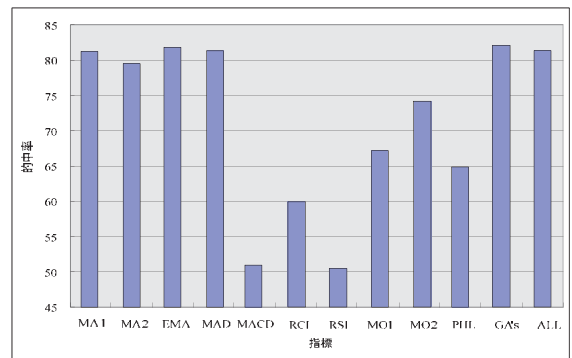


Fig. 4: 指標の違いによる的中率 (1993 年 IBM 株の例)

について予測を行った結果を Fig. 4 に示す。其々の指標の中身についてはあまり深く考えることなく、用意された道具の名前と考えて差し支えない。高速計算によって各時点ごとに一番使いそうな指標を探すという態度である。例えばこういう研究によって恒常的に良い成果を出ず指標が見つければ、それについて深く考え、より良い指標を考案する材料として使えるが、使ってみてあまり当たらない指標であれば捨て去って差し支えない。こういうことを計算で行うのが GA (遺伝アルゴリズム) である。

Fig. 4 の結果をみると GA's と書かれた棒グラフの的中率が最高となっており、これは GA によって選択された指標の組み合わせが最も的中率が高かったことを示すものであるが、MA1, EMA, MAD 等、一つの指標のみを用いた場合と比較して的中率にあまり大きな差は見られない。但し指標を組み合わせると単独で使うより一般的に的中率が上昇するのは確かなようで、実験に用いた全ての銘柄に対して同様の結果が得られた。

予測に有効な指標としてはテクニカル分析で広く用いられている移動平均 (MA1, MA2) や指数平滑移動平均 (EMA), 移動平均乖離率 (MAD) などの的中率が高く有効な指標であると言える。MACD や RSI といった指標はそれ単体では、ほとんど予測できないが、移動平均などと組み合わせると的中率が向上するという結果が得られた。

指標すべての組み合わせを用いて予測した場合の予測的中率はデータ数の多い銘柄ではそれなりの結果を得たが、データの少ない APC, BBY, SMRT 株では良い結果は得られなかった。全ての指標を組み合わせるとそれだけ多くの学習用データが必要となる。

指標の数を変えた場合、APC という株では指標数が 3 の場合に的中率が最も高く指標数を増やすほどの中率が低下する。一方、GM 株では指標数が 5 までは上昇するがそれ以上に増やすと却って低下する。一般に複数の指標を組み合わせると単独で使うより中率が向上するが、指標を多く使用しすぎると逆に低下してしまう傾

向がみられる。この理由としては、Fig. 2 に示すように、指標数が増えるにつれて戦略長 (L) がどんどん長くなるため、戦略数 (3^L) が爆発的に増えてしまい、最良の戦略を探しきれないままデータが尽きてしまうのが原因と考えられる。例えば単純移動平均 (MA1) のみを使用する時はその量が現在価格より高いか安いかの 2 通りしかないため 2 分木であるが、モメンタム (MO1) を併用すると 6 分木となり、探索空間が一挙に大きくなる。

このことを考慮すると、これまでに得られた結果は小手調べ段階とも言えるもので、まだまだ一般性に欠け、もっと多様なデータを大量に使う必要がある。

4. 価格のランダムウォーク仮説の検討

ところで、価格変動はランダムウォークであるという定説が長く存在した。それならば価格の予測をしようとする事自体が無意味だということになる。この定説は「市場は効率的である」という考えに基づけば当たり前なのだという。効率的とは、情報が素早く伝わることと同義である。仮に正しい価格というものがあるとして、それより今は安いということが解れば、その情報を素早く得た人は「安いうちに買おう」と考えて急いで買いに走るため、結果として買いが多く入って価格は上がってしまう。その結果として価格が正しい価格より上がりすぎたならば、その情報を素早く得た人は「高く売れるうちに早く売ってしまおう」と考えるので、売りが多く入って価格は下がる。その結果、価格は正しい価格の周囲を行ったり来たりするランダムウォークをするだけであるから、予測はできないことになる。

この考えはしかし問題を単純化しすぎている。情報が素早く伝わると言っても、伝達速度がゼロというわけではない。また、価格変動として、時間間隔を伝達速度よりずっと長く取れば確かに「効率的」かもしれないが、十分に情報が伝達されないほど短い時間間隔を取れば「効率的」になっているはずはなく、短期的な「安すぎる」価格や「高すぎる」価格を利用して（これを裁定機会と呼ぶ）濡れ手に粟の儲けを得ることも可能である。時間間隔を一日とした価格を「日次価格」と呼び、新聞やネットで日次価格の変化を誰でも知ることができる。一日という時間間隔が伝達速度よりずっと長いのかどうかは場合によるし、投資家の全てが同じ情報を得ているわけではないことを考えると、「効率的市場」という概念と、実際に一日という時間が情報伝達するのに十分長い時間間隔であるのかどうかとは直接には繋がらない。

マンテーニャとスタンレーによる 1995 年の論文 [8] では、S&P 500 指数の 5 分毎の価格変動がランダムウォークではなく分布の裾が広い（つまりグラフの横軸である価格変動の絶対値が正規分布に比べてずっと大きい値を

取る確率が高い）ことが実データを用いてわかりやすく表示されており、これに触発されて多くの研究がなされた。今ではもっと精密化されてはいるものの、裾の広い分布（fat-tail）については既に金融工学の常識となっており、膨大な文献が存在するが、著者が把握できる範囲を超えているため詳しくは金融工学の専門書を参照されたい。

筆者らは為替のティック価格と呼ばれる高頻度価格変動データを調べることにより、価格変化の取引毎（＝ティック毎）の条件付確率が長期間にわたって安定した数値を保ち、条件となる過去の価格変動としては 3 ティック前までは強い依存関係を示すが、4 ティック以上になると依存関係がぼやけてしまうことに気がついた [9]。この結果は 1995 年から 2001 年までの 5 年半にわたる日米為替の売り気配と買い気配のティック価格についてのものであり、本稿の 4 章以前の解析にもこれと同じデータ、および同時期のニューヨーク市場の株価約定価格を用いたため、予測も 3 ティック程度前までの情報が有効性を持つという暗黙の仮定が根底にある。こういった研究結果が一般性を持つためには、市場の仕組みがある程度安定していなければならないが、2010 年 1 月 4 日より東証に導入されたマイクロ秒単位の取引のできる arrowhead 制度により、それ以前は平均で 1 取引あたり 2-3 秒であったものが、1 秒あたり 20 万回以上の処理速度を持つようになったため、それまで少なくとも数秒以上の離散時間を単位に考えていた価格変動の論理関係が対応しないかもしれない世界の話になってしまった。こうなると、「前回価格がこう動いたから、次にはこうなるに違いない」という予測も成り立つかどうか怪しくなる。人間の考えをシミュレートする計算機のプログラムとしての人工知能ではなく、人が介在しない仮想世界の人工知能という対象を相手にどのような研究が可能なのか、悩ましいところである。

5. ゲーム理論的視点の重要性

ここまで述べてきた手法は予測者とそれを取り巻く環境である実価格データ時系列の間のゲームであるが、もっと一般的な視点からゲーム戦略の自動生成の問題を考察することも、それ自身興味のある研究対象である。そこで更にゲーム理論に近い形でプレーヤ対プレーヤが対等に「囚人のジレンマ」ゲームを行いつつ、単純で短い戦略から出発して、世代交代時における戦略の変異として、

- ① 点変異（遺伝子列中の 1 点が変わる）
- ② 2 倍変異（遺伝子列が 2 倍の長さにコピーされる）
- ③ 2 分裂変異（遺伝子列が半分に分れてどちらかになる）

の 3 種を併用して共進化させて行き、どのような戦略

が勝ちを収めるのか [10] を調べてゆくと面白いことが解った．問題を簡単化するためにプレイヤーの選択肢は 0 と 1 の 2 択とし，0 を裏切り，1 を協力に対応させ，戦略は 2 章の式 (2) のような 2 進数列で，Fig. 1 に類似した 2 分木の葉に対応させる．相手の直近の選択が 0 または 1 に対応して長さ 2 の戦略が 4 種類可能である．「00」という戦略は，相手の前の選択が何であろうと 0 (裏切り) 一辺倒を選択するものであり，「01」という戦略は相手の前の戦略を真似する「しっぺ返し」戦略となる．逆に「10」は相手の前の戦略の逆を選ぶ「逆しっぺ返し」，「11」は相手の前の戦略が何であろうと協力し続けるお人よし戦略である．このうち「01」が一般によく戦略とされている [11]．実際にシミュレーションを行ったところ，確かに「01」が勝ちを占める時代が 10000 世代に近い長期間続くのだが，より長い戦略である「1001」(パブロフ戦略) が現れるとこれが「01」より強いのでしばらく勝ちを占め，さらに数千世代を経て「0001」が現れるとこれの方が強いので勝ちを占めるが，10000 世代くらい後になると「0001」は「01」に負かされてしまい，1001→0001→01→1001→0001 という繰り返しが続くという現象が起きることがある．また場合によっては非常に長い戦略が勝ち進んで長さ 32 まで伸びることもある．この時は様々な戦略が現れるがその平均値を取ると，1001-0*0*0*0-0001，或いは更に特化した 1001-0001-0001-0001 の繰り返しが多く現れるという興味ある知見が得られた [12]．これらは「プレイヤー」対「プレイヤー」の戦略の進化であり，第 2 章の「プレイヤー」対「価格」の戦略と併せて，将来的には「プレイヤー」対「プレイヤー」対「価格」の 3 者ゲームへと広げてゆくことで現実と繋がってゆくものであり，その意味ではまだまだ玩具の段階と言わざるを得ない．

6. RMT を利用した主成分分析 (RMT-PCA)

これまででは，個々の株価の動きを追いながら数秒から数分先の動きを予測するというゲームシミュレーションの視点から価格予測を考えて来たが，別の視点から株式市場全体の動きを調べる一方法として，時系列間の相関行列の固有値問題を解く RMT-PCA 法について概観したい．

これは，市場で重要な動きをする株価時系列間で同時刻の内積を取って相関行列を作り，ランダム行列理論 (RMT) [13] を利用してランダム部分を捨象し，主成分を残すことで相互に強い連動をする株価群を抽出する方法である．RMT を利用した PCA であるので RMT-PCA と呼ぶ．ニューヨーク証券取引所の株価のうち S&P 500 に入っている取引量の多いものを選んで日次終値の変動値を扱った先行研究 [14] の手法を借りて，1 分毎の変動

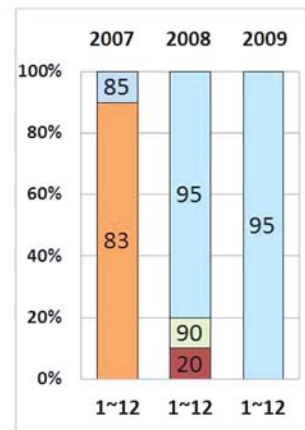


Fig. 5: 1 年間毎の主要業種の抽出結果

に対して有意な主成分を抽出し，これに基づいて株価相関の 4 半期や年次変動を追跡できる [15]．

ランダム行列理論によれば， N 種類の同時刻相関行列の固有値分布は，時系列の長さを T とすると， $N \rightarrow \infty$ ， $T \rightarrow \infty$ ， $Q = T/N =$ 定数の極限で

$$P_{\text{RMT}}(\lambda) = \frac{Q}{2\pi} \frac{\sqrt{(\lambda_+ - \lambda)(\lambda - \lambda_-)}}{\lambda} \quad (3)$$

となることが知られている．すなわち正規化されたランダム時系列の相関行列の固有値 λ は

$$\lambda_- < \lambda < \lambda_+ \quad (4)$$

の範囲に式 (3) に従って分布し，その上限と下限は

$$\lambda_{\pm} = 1 + \frac{1}{Q} \pm 2\sqrt{\frac{1}{Q}} \quad (5)$$

となる．

そこで得られた固有値のうちで，ランダム行列理論式に一致する，またはその範囲に存在するものをランダム成分とみなし，

$$\lambda \gg \lambda_+ \quad (6)$$

となる部分を降順に有意な主成分とする方法である．

この方法を用いて 2007 - 2009 年の東証株価の価格時系列の日次終値や日中データに適用した結果，確かにその時期に注目業種となった業種が抽出された [16]．

RMT-PCA を 1 年ごとの価格時系列に適用して抽出した東京証券取引所の注目業種を Fig. 5 に示す．3 本の棒グラフは左から 2007 年，2008 年，2009 年の其々一年間の株価時系列間の相関行列の固有値問題を解いて求めた，其々の時期に強い相関を示す業種の内訳である．ここでは最大固有値に対応する主成分の固有ベクトル成分には業種の特徴が出ず市場全体の動きに連動するものであるため，第 2 固有値に対応する固有ベクトル成分の分布を正值部分と負値部分の絶対値のそれぞれに分けて降順に 10 成分を選び，その株価が属する業種の比率を棒

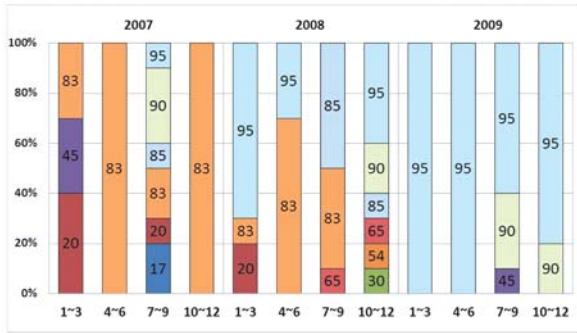


Fig. 6: 3ヶ月毎の主要業種の抽出結果

グラフに表す．第4固有値以上の固有ベクトル成分はほぼランダムであるため明瞭な特徴を持たず，第3固有値に対応する主成分の固有ベクトル成分には特徴はあるものの精度は第2固有ベクトルより弱いため，ここでは第2固有ベクトル成分の正負それぞれの最大10成分を選んだ．

2007年の大半が83であり一部が85となっているのは，東証株価の銘柄コードが8300番台である株と8500番台となる株に其々強い相関が見られたことを示す．これらは銀行や証券会社からなる金融株である．相関が強い業種はそれに属する複数株が一斉に同方向に動き，共振現象を起こしていると考えられることができる．

2009年は棒グラフ全体が95で占められており，これは9500番台の電力・ガス等のエネルギー供給株が一斉に共振状態になり，大変目立った動きをする業種に躍り出たことを示す．これは実際に東証株価の主体がリーマンショックを境に金融株から電力・ガス等に移行したことに対応している．

この結果に自信を得て，RMT-PCA手法を東証株価の3ヵ月単位のデータで解析すると，もっと細かい年内遷移を追えるはずである．その結果 [17] は，Fig. 6 に示すように2007年秋のサブ・プライム・ローンの破綻と2008年夏のリーマン・ブラザーズの倒産の影響と見られる市場の混乱を見事に捉えていることがわかる．

これらの解析で対象としたデータは時系列そのままではなく，同時刻相関を取れるように成形している，日中変動の場合，1分毎の価格変化が必ずしも全株価について行われているわけではないので，欠損部分は前の価格が動かなかったと考えて1つ前の数値と同じとした．ただしこのような欠損部分が時系列長 T の2割以上となるような株は使用しないこととした．この理由により，時期ごとに異なる株価数 N のデータを扱うことになる．そのため，データ毎にパラメータ Q の値も異なるが，ここで使用したデータは全て本手法の適用条件である $Q > 1$ と $N > 100$ [14–19] を満たしていることが Table 2 に示されている．

Table 2: 等間隔に成形後の3ヵ月毎，および1年毎のデータ [17]

| Year | Quarter | T | N | Q=T/N | λ_+ | # $\lambda > \lambda_+$ |
|------|---------|------|-----|-------|-------------|-------------------------|
| 2007 | I | 642 | 486 | 1.32 | 3.50 | 13 |
| | II | 681 | 486 | 1.40 | 3.40 | 22 |
| | III | 681 | 489 | 1.39 | 3.41 | 18 |
| | IV | 675 | 492 | 1.37 | 3.44 | 14 |
| 2008 | I | 642 | 488 | 1.32 | 3.50 | 11 |
| | II | 681 | 491 | 1.39 | 3.42 | 14 |
| | III | 692 | 492 | 1.41 | 3.40 | 15 |
| | IV | 664 | 487 | 1.36 | 3.45 | 13 |
| 2009 | I | 642 | 490 | 1.31 | 3.51 | 11 |
| | II | 659 | 486 | 1.36 | 3.46 | 13 |
| | III | 681 | 485 | 1.40 | 3.40 | 13 |
| | IV | 670 | 483 | 1.39 | 3.42 | 15 |
| 2007 | all | 2682 | 477 | 5.62 | 2.02 | 20 |
| 2008 | all | 2682 | 480 | 5.59 | 2.03 | 19 |
| 2009 | all | 2655 | 476 | 5.58 | 2.03 | 16 |

7. RMT を利用した乱数度計測法: RMT-test

一つの時系列から長さ T の等長列を N 個採取し，この間の相関行列の固有値分布のヒストグラムと式 (3) の比較により目視で乱数度を判断するのが RMT テストの定性評価 [18–20] である．Fig. 7(左) に示す例のように，角のある線で示すヒストグラムと，曲線で示す RMT 理論式 (データと同じ Q に対応する式 (3) のグラフ) の両者が一致すれば RMT テストに合格したとし，Fig. 7(右) に示す例のように不一致ならば RMT テストに不合格と判断する．この定性評価は直感的に見やすく，形状の特徴を把握しやすい利点がある．

一方，定性評価だけでは，乱数度の非常に近いものどうしを比較しにくいのでモーメントを取って定量化する．定量化の具体的な手続きは次のようになる．相関行列 C の固有値分布の k 次モーメントは固有値の k 乗の平均値であり，次式で求められる．

$$m_k = \frac{1}{N} \sum_{i=1}^N \lambda_i^k \tag{7}$$

対応する理論値は固有値の理論分布式 (3) から以下のよう求められる．

$$\mu_k = E(\lambda^k) = \int_{\lambda_-}^{\lambda_+} \lambda^k P_{\text{RMT}}(\lambda) d\lambda \tag{8}$$

両者の誤差の逆数が乱数度となるが，当面，以下に定義する誤差

$$\text{誤差} = m_k / \mu_k - 1 \tag{9}$$

を用いて比較を行う．最大値付近の誤差が大きく反映する $k = 6$ の場合が最も比較しやすいことから各データの

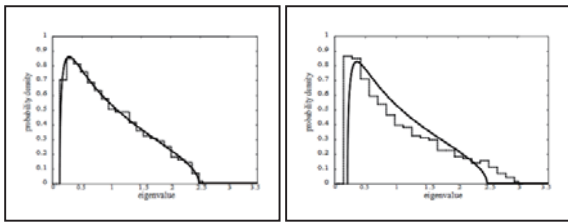


Fig. 7: RMT テスト定性評価の例 (左: 合格, 右: 不合格)

Table 3: 2007 年の東証 1 分足から求めた乱数度順位

| 順位 | 業種 | 企業コード | 誤差絶対値 |
|-----|-------|-------|--------|
| 1 | 電力・ガス | 9504 | 26.4 |
| 2 | 機械 | 6460 | 37.6 |
| 3 | 電力・ガス | 9506 | 38.2 |
| 4 | 電力・ガス | 9508 | 43.3 |
| 5 | 情報・通信 | 4676 | 44.9 |
| ... | | | |
| 207 | 電気機器 | 6506 | 740.9 |
| 208 | 非鉄金属 | 5802 | 797.3 |
| 209 | 化学 | 4043 | 799.8 |
| 210 | 鉄鋼 | 5541 | 1001.5 |
| 211 | 輸送用機器 | 7201 | 1209.6 |

6 次モーメントの誤差絶対値を RMT テストの定量評価結果とし、それにより株価変動のランダム性を数値化することで、“6 次モーメントの誤差が小さい方が乱数度が高い”と定義する [20, 21]。

8. 株価ゆらぎの乱数度とその後の株価安全性

2007 年から 2009 年の東証 TOPIX500 の 3 年間 1 分毎の価格データを対象として分析した結果、前年に乱数度が高い方が翌年の収益率が高いという観測結果が得られた。この結果は Table 3 から、2007 年 1 年間に最も乱数度が高い株価 (9504) の収益率が、その翌年の 2008 年に日経平均の収益率より一貫して高く、乱数度最低の株価 (7201) の収益率は日経平均に比較して一貫して低いという形で Fig. 8 に示される。

同様の結果は Table 4 に於いて 2009 年の株価の乱数度最高と最低の株の翌年 2010 年における収益率についても成立することが Fig. 9 により示される。

ところが 2008 年の乱数度とその翌年 2009 年の収益率に関してはこの法則が成立しない。実際、Table 5 に於いて高乱数度を示す株価は安全株ではない。これは

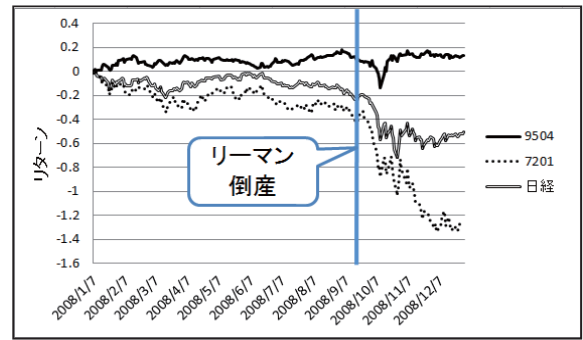


Fig. 8: 前年の 2007 年に於いて乱数度最高の株 (9504) の翌年収益は日経平均株価より安全であり、日経平均は乱数度最低株 (7201) より安全である

Table 4: 2009 年の東証 1 分足から求めた乱数度順位

| 順位 | 業種 | 企業コード | 誤差絶対値の平均値 |
|-----|-------|-------|-----------|
| 1 | 電力・ガス | 9508 | 25.5 |
| 2 | 電力・ガス | 9509 | 28.1 |
| 3 | 電力・ガス | 9506 | 31.6 |
| 4 | 電力・ガス | 9502 | 33.2 |
| 5 | 小売業 | 2651 | 36.4 |
| ... | | | |
| 225 | 非鉄金属 | 5713 | 1039.5 |
| 226 | 電気機器 | 4902 | 1072.4 |
| 227 | 機械 | 6301 | 1090.9 |
| 228 | 鉄鋼 | 5541 | 1128.2 |
| 229 | 卸売業 | 8058 | 1249.5 |

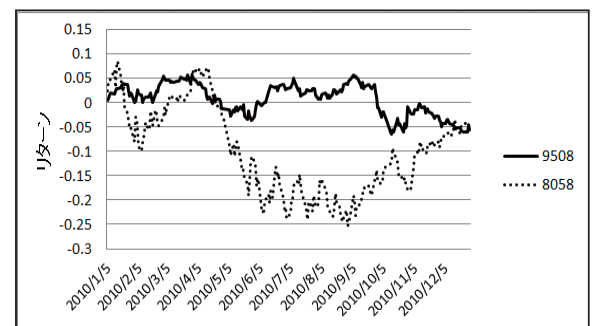


Fig. 9: 前年の 2009 年に乱数度最高の株 9508 が最低株 8058 より 2010 年に於いて安全性が高い

しかし、2008 年 10 月に起きたリーマンショックの影響と考えると解決する。乱数度を計算するデータとして、ショック以前のものだけ、すなわち 2008 年については 1 年分ではなく、1 月から 8 月までの価格揺らぎから乱数

Table 5: 2008 年の東証 1 分足から求めた乱数度順位

| 順位 | 業種 | 企業コード | 誤差絶対値 |
|-----|-------|-------|--------|
| 1 | 銀行業 | 8308 | 28.4 |
| 2 | 機械 | 7004 | 30.9 |
| 3 | 輸送用機器 | 7211 | 31.8 |
| 4 | 電力・ガス | 9502 | 32.3 |
| 5 | 電力・ガス | 9508 | 36.3 |
| ... | | | |
| 236 | 電気機器 | 4902 | 1604.0 |
| 237 | 電気機器 | 6954 | 1611.2 |
| 238 | 電気機器 | 7752 | 1646.9 |
| 239 | 証券業 | 8604 | 2059.9 |
| 240 | 海運業 | 9104 | 2097.5 |

Table 6: 2008 年 (Jan. – Aug.) のデータによる乱数度順位

| 順位 | 業種 | 企業コード | 誤差絶対値の 平均値 |
|-----|-------|-------|---------------|
| 1 | 電力・ガス | 9506 | 11.6 |
| 2 | 電気機器 | 6728 | 13.1 |
| 3 | 食料品 | 2267 | 19.1 |
| 4 | 電力・ガス | 9502 | 19.1 |
| 5 | 小売業 | 2685 | 22.2 |
| ... | | | |
| 230 | 電気機器 | 6762 | 183.0 |
| 231 | 電気機器 | 6503 | 183.7 |
| 232 | 銀行業 | 8306 | 190.0 |
| 233 | 証券業 | 8604 | 222.8 |
| 234 | 輸送用機器 | 7201 | 257.9 |

度を測ると、Table 6 に示すように安全株が乱数度上位となって高乱数度の株が低乱数度の株より翌年の収益率が良いと言う結果が Fig. 10 に見るようにこの年に対しても成立する [22–24]。しかしここで示す乱数度と収益率の関係は、対象とした価格時系列が 2007 年から 2009 年にかけての「下降相場」であるため限定的な結果に留まる。「上昇相場」も併せて調べることによりもっと一般的な結果を得る必要があり今後の課題としたい。

9. まとめと今後の課題

本稿ではゲーム理論による戦略の自動生成と RMT-PCA、および前年度の価格ゆらぎの乱数度を利用する方

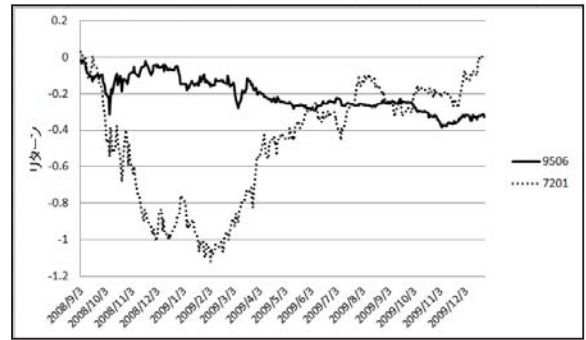


Fig. 10: 前年に乱数度最高の 9506 が乱数度最低の 8604 より翌年の対数収益が高く安全であることを示す

法の 3 つの異なる手法を紹介した。

結果は一部の時期にしか当てはまらない過渡的なものであり、3 手法はそれぞれ価格ゆらぎを別の側面から見たものに相当するが、ここで検討した手法はそれぞれ独立に正当性と重要性を持つことを示せたと思う。

結果として、価格の予測はある程度まで可能であり、そこで用いた手法は更に掘り下げて検討する価値のあるものである。

単一時系列として見た場合の価格ゆらぎは、ゲーム論的な考えにより精度は低いが数歩先の価格の変化に対して何らかの予測はできる。また、もっと長期にわたる価格の動向の予測は、ランダム性を利用して主成分抽出を行う RMT-PCA や、その手法を転用した乱数度計測法 RMT-test を利用した乱数度を測定することでその傾向をある程度予測できる。

今後の課題は、より広範囲の種類・期間に対するデータに於いてここで検討した手法の有効性を検証することであり、そのためにはデータの収集・保管・公開のルールの急速な整備が必要であると考えられる。

参考文献

- [1] J. von Neumann and Oskar Morgenstern: “Theory of Games and Economic Behavior,” Princeton University Press, Princeton, 1953.
邦訳: J. フォン・ノイマン, O. モルゲンシュテルン: ゲームの理論と経済行動 I, II, III, 銀林浩, 橋本和美, 宮本敏雄監訳, 阿部修一, 橋本和美訳, 筑摩書房, 2009.
- [2] M. Tanaka-Yamawaki and T. Motoyama: “Predicting the Tick-wise Price Fluctuations by Means of Evolutionary Computation,” IEEE-CEC2004, pp. 955-958, 2004.
- [3] NTAA: “Nihon Technical Bunseki Taizen,” Nihon Keizai Shimbun, Inc., 2004, (in Japanese), Early version translated to English, NTAA: “Analysis of Stock Prices in Japan,” 1988.
- [4] M. Tanaka-Yamawaki and S. Tokuoka: Physica A, 383, pp. 125-133, 2007.

- [5] 田中美栄子, 徳岡聖二: 情報処理学会論文誌数理モデル化と応用, Vol.49(SIG4), pp. 88-91, 2008.
- [6] NYSE Trade and Quotes: <http://www.nysedata.com>
- [7] 徳岡聖二, 田中美栄子: 情報処理学会論文誌数理モデル化と応用, Vol.48(SIG19), pp. 68-74, 2007.
- [8] R. N. Mantegna and H. E. Stanley, "Scaling behaviour in the dynamics of an economic index," *Nature*, Vol.376, pp. 46-49, 1995.
- [9] M. Tanaka-Yamawaki: "Stability of Markovian Structure Observed in High Frequency Foreign Exchange Data," *Annals of Institute for Statistical Mathematics (AISM)*, Vol.55, pp. 437-446, 2003.
- [10] R. Axelrod: "The Evolution of Cooperation," 1984, 松田耕治 (訳): つきあい方の科学, ミネルヴァ書房, 1998.
- [11] K. Lindgren: "Evolutionary Phenomena in Simple Dynamics," Addison-Wesley, *Artificial Life II*, pp. 295-312, 1990.
- [12] 糸井良太, 田中美栄子: ノイズ付き進化型繰返し凶人のジレンマにおける長寿戦略の探究, 情報処理学会論文誌数理モデル化と応用, Vol.6, No.1, pp. 31-37, 2013.
- [13] M. L. Mehta: "Random Matrices," 3rd (Ed.), Academic Press, San Diego, 2004.
- [14] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley: "Random matrix approach to cross correlations in financial data," *Physical Review E*, p. 066126, 2002.
- [15] M. Tanaka-Yamawaki: "Cross correlation of intra-day stock prices in comparison to Random Matrix Theory," *Intelligent Information Management*, Vol.3, pp. 65-70, 2011.
- [16] 木戸丈剛, 楊欣, 田中美栄子, 高石哲弥: ランダム行列理論を用いた主成分抽出法による日本と米国の株式市場における主要セクタの変遷, 情報処理学会論文誌数理モデル化と応用, Vol.4, No.4, pp. 104-110, 2011.
- [17] M. Tanaka-Yamawaki, X. Yang, T. Kido, and A. Yamamoto: "Extracting Market Trends from the Cross Correlation between Stock Time Series," J. W. Tweedale and L. C. Jain (Eds.), Springer, Heidelberg, KES 2012, CCIS, Vol.246, pp. 25-38, 2013.
- [18] 田中美栄子, 糸井良太, 楊欣: ランダム行列理論を用いた乱数度評価法の提案, 情報処理学会論文誌数理モデル化と応用, Vol.5, No.1, pp. 1-8, 2012.
- [19] X. Yang, R. Itoi, and M. Tanaka-Yamawaki: "Testing Randomness by Means of RMT Formula," *SIST, Intelligent Decision Technologies*, Vol.10, pp. 589-596, 2011.
- [20] X. Yang, R. Itoi and M. Tanaka-Yamawaki: "Testing randomness by means of Random Matrix Theory," *Progress of Theoretical Physics Supplement*, Vol.194, pp. 73-83, 2012.
- [21] M. Tanaka-Yamawaki, X. Yang, and R. Itoi: "Moment Approach for quantitative evaluation of randomness based on RMT formula," *SIST, Intelligent Decision Technologies*, Vol.16, pp. 423-432, 2012.
- [22] X. Yang, Y. Mikamori, and M. Tanaka-Yamawaki: "Application of the RMT-test on Real Data-Hash Function and Tick Data of Stock Prices," *JAFEE 2013*, March 16-17, Chuo University, 2013.
- [23] M. Tanaka-Yamawaki, X. Yang, Y. Mikamori: "Randomness as the Security Levels of Investments," R-Neves-Silva, et.al. (Eds.), IOS Press, *Intelligent Decision Technologies (Proc. 5th KES-IDT)*, pp. 344-352, 2013, doi/978-1-61499-264-6-344.
- [24] X. Yang and M. Tanaka-Yamawaki: "A study on the randomness of stock prices by using the RMT-test," *The 12th Asia Pacific Physics Conf. (APPC2013)* July 14-19, 2013, Makuhari Chiba, Japan. (To be published in *JPSJ Supplement*)

田中 美栄子



1950年8月4日生. 1974年京都大学理学部卒. 1983年ロチェスター大学物理天文学大学院 Ph.D. コース修了. 2003年鳥取大学工学部教授, 現在に至る. 複雑系科学, 経済物理学, 計算知能などの研究に従事. Ph.D. in physics. 日本物理学会, 情報処理学会などの会員.

楊 欣



1984年生. 2007年中国吉林大学工学部物流工科学科卒業. 2009年中国吉林大学大学院工学研究科物流工学専攻修士課程修了. 2013年9月鳥取大学大学院工学研究科博士課程修了予定.

糸井 良太



1989年生. 2011年鳥取大学工学部卒. 2013年鳥取大学工学研究科情報エレクトロニクス専攻修了. 工学修士. 2013年4月より(株)シンフォームに勤務.
